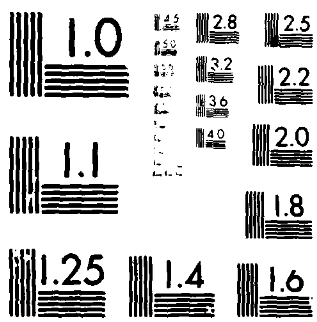
AD-A085 211 ARMY MILITARY PERSONNEL CENTER ALEXANDRIA VA F/G 5/10
THE NOVELTY OF HUMAN SELF-ASSESSMENT: IMPLICATIONS FOR LEARNING--ETC(U)
APR 80 R W CROUSE

UNCLASSIFIED NL

1 OF 1
AD
A085211

END
DATE
FILMED
6-80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

THE NOVELTY OF HUMAN SELF-ASSESSMENT: IMPLICATIONS FOR LEARNING AND TRAINING

BY

ROBIN WESLEY CROUSE JR., B.S.

A Thesis submitted to the Graduate School
in partial fulfillment of the requirements
for the Degree
Master of Arts

Major Subject: Psychology (Experimental)

New Mexico State University
Las Cruces, New Mexico
April 1980

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. AD-A085211

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6 The Novelty of Human Self-Assessment: Implications for Learning and Training.

5. TYPE OF REPORT & PERIOD COVERED
9 Final Report.
~~29 April 1980~~

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
10 Robin Wesley Crouse, Jr.
CPT, IN

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Student, HQDA, MILPERCEN (DAPC-OPP-E)
200 Stovall Street
Alexandria, VA 22332

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
HQDA, MILPERCEN, ATTN: DAPC-CPP-E
200 Stovall Street
Alexandria, VA 22332

12. REPORT DATE
11 29 Apr[redacted]80

13. NUMBER OF PAGES
80

14. MONITORING AGENCY NAME & ADDRESS (*if different from Controlling Office*)
12 84

15. SECURITY CLASS. (*of this report*)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)
Approved for public release, distribution unlimited

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES
Document is a Thesis completed as a requirement for the Master of Arts Degree in Experimental Psychology at New Mexico State University.

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)
Learning facilitation thru overt self-assessment procedures.
Cortical arousal thru confirmation/disconfirmation of outcome expectancies.

Paired-associates learning paradigm.

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)
Overt self-assessment appears to induce an increased use of "sure" assessments(over-confidence), which leads to both greater disconfirmation and greater confirmation of subject-held expectancies Outcomes that disconfirm or confirm expectancies are "biological" events that are "novel" or "reinforcing", respectively, and that have the capacity of eliciting the fundamental condition necessary for learning, which is cortical arousal.

Theory is examined using paired-associates learning paradigm.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

391191

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

INSTRUCTIONS FOR PREPARATION OF REPORT DOCUMENTATION PAGE

RESPONSIBILITY. The controlling DoD office will be responsible for completion of the Report Documentation Page, DD Form 1473, in all technical reports prepared by or for DoD organizations.

CLASSIFICATION. Since this Report Documentation Page, DD Form 1473, is used in preparing announcements, bibliographies, and data banks, it should be unclassified if possible. If a classification is required, identify the classified items on the page by the appropriate symbol.

COMPLETION GUIDE

General. Make Blocks 1, 4, 5, 6, 7, 11, 13, 15, and 16 agree with the corresponding information on the report cover. Leave Blocks 2 and 3 blank.

Block 1. Report Number. Enter the unique alphanumeric report number shown on the cover.

Block 2. Government Accession No. Leave Blank. This space is for use by the Defense Documentation Center.

Block 3. Recipient's Catalog Number. Leave blank. This space is for the use of the report recipient to assist in future retrieval of the document.

Block 4. Title and Subtitle. Enter the title in all capital letters exactly as it appears on the publication. Titles should be unclassified whenever possible. Write out the English equivalent for Greek letters and mathematical symbols in the title *(see "Abstracting Scientific and Technical Reports of Defense-sponsored RDT/E," AD-667 000).* If the report has a subtitle, this subtitle should follow the main title, be separated by a comma or semicolon if appropriate, and be initially capitalized. If a publication has a title in a foreign language, translate the title into English and follow the English translation with the title in the original language. Make every effort to simplify the title before publication.

Block 5. Type of Report and Period Covered. Indicate here whether report is interim, final, etc., and, if applicable, inclusive dates of period covered, such as the life of a contract covered in a final contractor report.

Block 6. Performing Organization Report Number. Only numbers other than the official report number shown in Block 1, such as series numbers for in-house reports or a contractor/grantee number assigned by him, will be placed in this space. If no such numbers are used, leave this space blank.

Block 7. Author*(s)*. *Include corresponding information from the report cover.* Give the name*(s)* of the author*(s)* in conventional order *(for example, John R. Doe or, if author prefers, J. Robert Doe).* In addition, list the affiliation of an author if it differs from that of the performing organization.

Block 8. Contract or Grant Number*(s)*. For a contractor or grantee report, enter the complete contract or grant number*(s)* under which the work reported was accomplished. Leave blank in in-house reports.

Block 9. Performing Organization Name and Address. For in-house reports enter the name and address, including office symbol, of the performing activity. For contractor or grantee reports enter the name and address of the contractor or grantee who prepared the report and identify the appropriate *corporate division, school, laboratory, etc., of the author.* List city, state, and ZIP Code.

Block 10. Program Element, Project, Task Area, and Work Unit Numbers. Enter here the number code from the applicable Department of Defense form, such as the DD Form 1498, "Research and Technology Work Unit Summary" or the DD Form 1634. "Research and Development Planning Summary," which identifies the program element, project, task area, and work unit or equivalent under which the work was authorized.

Block 11. Controlling Office Name and Address. Enter the full, official name and address, including office symbol, of the controlling office. *(Equates to funding/sponsoring agency. For definition see DoD Directive 5200.20, "Distribution Statements on Technical Documents.")*

Block 12. Report Date. Enter here the day, month, and year or month and year as shown on the cover.

Block 13. Number of Pages. Enter the total number of pages.

Block 14. Monitoring Agency Name and Address *(if different from Controlling Office).* For use when the controlling or funding office does not directly administer a project, contract, or grant, but delegates the administrative responsibility to another organization.

Blocks 15 & 15a. Security Classification of the Report: Declassification/Downgrading Schedule of the Report. Enter in 15 the highest classification of the report. If appropriate, enter in 15a the declassification/downgrading schedule of the report, using the abbreviations for declassification/downgrading schedules listed in paragraph 4-207 of DoD 5200.1-R.

Block 16. Distribution Statement of the Report. Insert here the applicable distribution statement of the report from DoD Directive 5200.20, "Distribution Statements on Technical Documents."

Block 17. Distribution Statement *(of the abstract entered in Block 20, if different from the distribution statement of the report).* Insert here the applicable distribution statement of the abstract from DoD Directive 5200.20, "Distribution Statements on Technical Documents."

Block 18. Supplementary Notes. Enter information not included elsewhere but useful, such as: Prepared in cooperation with . . . Translation of *(or by)* . . . Presented at conference of . . . To be published in . . .

Block 19. Key Words. Select terms or short phrases that identify the principal subjects covered in the report, and are sufficiently specific and precise to be used as index entries for cataloging, conforming to standard terminology. The DoD "Thesaurus of Engineering and Scientific Terms" *(TEST)*, AD-672 000, can be helpful.

Block 20. Abstract. The abstract should be a brief *(not to exceed 200 words)* factual summary of the most significant information contained in the report. *If possible, the abstract of a classified report should be unclassified* and the abstract to an unclassified report should consist of publicly- releasable information. If the report contains a significant bibliography or literature survey, mention it here. For information on preparing abstracts see "Abstracting Scientific and Technical Reports of Defense-Sponsored RDT&E," AD-667 000.

★ U.S. GPO: 1974—540-847/9082

THE NOVELTY OF HUMAN SELF-ASSESSMENT: IMPLICATIONS FOR LEARNING AND TRAINING

CAPTAIN ROBIN WESLEY CROUSE JR.
HQDA, MILPERCEN (DAPC-OPP-E)
200 Stovall Street
Alexandria, VA 22332

Final Report
29 April 1980

A Thesis submitted to the Graduate School,
New Mexico State University, Las Cruces, New Mexico,
in partial fulfillment of the requirements for the degree
Master of Arts.

THE NOVELTY OF HUMAN SELF-ASSESSMENT: IMPLICATIONS FOR LEARNING AND TRAINING

BY

ROBIN WESLEY CROUSE JR., B.S.

A Thesis submitted to the Graduate School
in partial fulfillment of the requirements
for the Degree
Master of Arts

Major Subject: Psychology (Experimental)

New Mexico State University
Las Cruces, New Mexico
April 1980

"The Novelty of Human Self-Assessment: Implications for Learning and Training," a thesis prepared by Robin Wesley Crouse Jr. in partial fulfillment of the requirements for the degree, Master of Arts, has been approved and accepted by the following:

W.M. Matchett
Dean of the Graduate School

Darwin P Hunt

Darwin P. Hunt
Chairman of the Examining Committee

29 April 1980

Date

Committee in Charge:

Dr. Darwin P. Hunt, Chairman
Dr. James V. Bradley
Dr. Richard E. Christ
Dr. Roger W. Schvaneveldt

A

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. Darwin P. Hunt for the use of his 1978 data and his laboratory, and for his continuing and valuable guidance throughout the conduct of this experiment.

The author wishes to thank Mr. Al Moreno, the department technician, for his expert and patient assistance in times of mechanical distress.

The author wishes to thank Dr. James V. Bradley, Dr. Roger W. Schvaneveldt, and Dr. Richard E. Christ for their time spent and guidance provided.

Finally, the author wishes to thank his wife, Shelly, for her patience and support.

VITA

June 9, 1949 - Born at Roswell, New Mexico

1967 - Graduated from Roswell Senior High School

1971 - B.S. in Geography, New Mexico State University

1971-Current - Commissioned Officer, U.S. Army

ABSTRACT

THE NOVELTY OF HUMAN SELF-ASSESSMENT:
IMPLICATIONS FOR LEARNING AND TRAINING

BY

ROBIN WESLEY CROUSE JR., B.S.

Master of Arts in Psychology
New Mexico State University
Las Cruces, New Mexico
Dr. Darwin P. Hunt, Chairman

A report by Hunt (1978) is used as a basis for development of a postulated explanation for the learning facilitation achieved by an overt self-assessment process. Hunt (1978) found that the addition of an overt self-assessment step to the stimulus-response cycle of a paired-associates learning task facilitates learning, as much as 25% over a normal learning control condition. An item by item re-analysis of Hunt's (1978) data, by this student, shows that the response-assessment order of responding produces more extensive use of "sure" assessments, than does the assessment-response order of responding. It is herein proposed that overt

self-assessment induces an increased use of "sure" assessments, which leads to both greater disconfirmation and greater confirmation of subject-held expectancies of assessment and response outcomes. Outcomes that disconfirm or confirm expectancies are "biological" events that are either "novel" or "reinforcing" in nature, and that have the capacity of eliciting the fundamental condition necessary for learning, which is cortical arousal (Johnston, 1979). A paired-associates learning experiment using CVC trigrams of a 65% meaningfulness level sought to replicate portions of Hunt's (1978) study for comparison with the effects of a new variable, nature of the assessment scale. The results of the experiment were negative. However, a cross-comparison of the data and outcome of the Hunt (1978) study, with the data and outcome of this study shows an internal consistency with the theoretical notions offered in this report.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDIX TABLES

LIST OF FIGURES

Introduction

Training techniques or procedures that facilitate the acquisition of knowledge or skills by trainees in a training program can save time and money. Such learning facilitators can reduce the drain on finite resources in both direct and indirect ways. If a given technique facilitates learning of a required skill so that trainees acquire the skill more rapidly and/or retain it better, then less training time or less subsequent refresher training will be required. Directly an advantage will have been achieved by such facilitation in the training of the skill. Also, an indirect advantage will accrue from the freeing of training assets once required but no longer necessary to teach the skill. Also, if the skill is a portion of a skill hierarchy or one of a group of skills, additional advantage should accrue by transfer or association.

A facilitatator, as described, could be any method, technique, procedure, or apparatus that somehow improves learning efficiency in individuals or groups. By its broad definition it could be skill-dependent, or generally effective and virtually skill-independent. Certainly, any advantage would be welcome, but a generalizable advantage is most desirable.

In search of a broadly effective learning facilitator it would seem appropriate to identify and examine those things which are common to all human training situations.

The single most common factor is the trainee and his/her cognitive processes. Therefore, it seems reasonable that the domain of cognitive processing might be a fruitful area in which to search for a general learning facilitator.

Hunt (1978) employs a procedure referred to as the Human Self-Assessment Process, which requires an individual to overtly assess his/her level of sureness that a decision he/she has derived or a response he/she has made is correct. In essence, Hunt sought to examine whether or not the addition of a self-assessment step to the stimulus-response cycle of a paired-associates learning (PAL) task would affect learning efficiency in a favorable way.

He asked subjects to learn the names of eight different types of hand pliers, by matching the stimulus images of the pliers with their names. Subjects were processed one at a time at a computer (PDP/8e) controlled keyboard, and each subject was exposed to only one of the nine possible treatments. Number of response steps, number of self-assessment sureness levels, and order of response were varied to create the nine treatments, as shown in Table 1. The criterion for learning was established as the errorless completion of two sequential trials of eight stimuli each. The dependent variable was the number of trials to criterion.

Table 1

Treatments of Hunt (1978) Paradigm

Treatment[a]	Sequence Stimulus-Response[b]	Mean Trials-to-Criterion[c]
M	S - R - KR	20.5
MX	S - R - X - KR	17
MK2	S - R - K2 - KR	17.5
MK4	S - R - K4 - KR	16
MK8	S - R - K8 - KR	15
XM	S - X - R - KR	16.5
K2M	S - K2 - R - KR	21
K4M	S - K4 - R - KR	18
K8M	S - K8 - R - KR	17.5

a: Treatment symbology is extracted from a model employed by Hunt (1978, page 4).

b. Key

S: Stimulus K#: Assessment, # levels

R: Response KR: Knowledge-of-Results

X: Motor Component; Used in control treatments to match the order and number of motor responses in assessment treatments.

c: Means shown are extrapolations taken from a figure used by Hunt (1978, page 34).

Generally, it was found that the number of trials to criterion was less for most of the experimental treatments than it was for the control treatment (M). The inclusion of an assessment step in the response process does appear to improve learning efficiency. Specific findings of interest to this discussion are shown below.

a. MK treatments expedited learning more than did the KM treatments.

b. Treatments with more levels of sureness (MK8 and K8M) expedited learning more than did those with few levels of sureness.

c. The MK8 treatment produced the best learning efficiency, requiring approximately 25% fewer trials than the normal learning (M) condition to reach criterion.

The potential contribution of the Human Self Assessment Process, as described by Hunt (1978), is significant both in practical and theoretical ways. A 25% facilitation in learning can not easily be ignored, yet it is too early to broadly apply the process as a teaching method. The essence of the process needs to be teased out, and the domain over which it might apply needs be defined. An obvious and basic question is "why?", or rather "how?" does the self assessment process achieve its learning facilitation effect. Equally intriguing is the differential performance of the MK and KM treatments. Why is one assessment order more

effective than the other? It is to these ends that this paper is directed.

There are two notions of particular interest in Hunt's (1978) description of the assessment process: first, that the human operator employs an internal representational system to model external events, and second, that the human operator employs a covert assessment mechanism to control the selection of responses for overt execution. Together these ideas form the crux of the theoretical framework, which is aimed at describing the human operator's covert decisional processes. The following quote from Hunt (1978) captures the essence of these notions:

> the performance of an individual....importantly depends upon the validity and reliability with which the person can assess whether items of knowledge and responses which are relevant to the performance of the task are stored in his/her own memory, are retrievable from it and are executable. (page 1)

Typically, as an individual is confronted with a situation (stimulus), cognitive modeling operations will select a tentative response and simulate its execution. The simulation procedure generates a covert sureness, which must exceed a situationally determined criterion sureness in order for the response to be released for actual execution. It is herein assumed that all human operators employ a set of covert processes approximately like that described above. The Human Self-Assessment Process (Hunt, 1978) is an overt manipulation of these postulated covert events.

The construct of cognitive modeling operations (CMO) finds a good deal of support in pertinent literature. Attneave (1974) describes the internal representational system as being capable of simplistic representation of tri-dimensional analogue imagery. Modeling operations preserve parameters, functional relationships, and applicable rules of the real-world situation being modeled. Using language as its basis of "knowing", CMO relies upon a descriptive system, memory, as a storage method. Applicably, an individual, confronted with some sort of decisional situation in reality, creates a covert analogue situation upon which plausible solutions are attempted and results observed. A successful analogue solution can then be applied in reality to actually resolve the external real-world situation.

A feature of CMO that is key to this discussion is predictive (assessment) capacity. As Attneave (1974, page 498) and Kelly (1968, page 7) point out, the human operator can simulate memory events as well as current input, and perform vicarious manipulations and locomotion without regard to space-time limitations. Employment of CMO over time allows the individual to develop plans for behavior (Miller, Gallanter, and Pribram, 1960), which consist of decisions, and predictions or estimates about the actual consequences of the decisions. In terms of the assessment process in a paired-associates learning (PAL) task, the

decision is the selection of a response, and the prediction is the assessment of the correctness of the response.

In our daily lives we have frequent occasion to make decisions about events at points in the future. When an event is initially considered, we derive a particular decision or set of decisions (plan), and consequently generate some degree of sureness (prediction) that the plan is appropriate and that it will have the effect we wish. As the event grows closer in time, new information that becomes available may cause us to alter the plan or our sureness about it. Additionally, we may have cause to elaborate our plan to others and overtly assess its appropriateness and predicted effect. This overt elaboration and assessment procedure might also generate changes in the plan or our sureness about it. When the time comes for the execution of the plan (response), there may be on-the-spot changes of the plan, which may necessitate an adjustment of sureness. Once the plan has been executed it is likely that we will consider if the plan was executed as intended. This verification process will generate a sureness level, which can be compared to the sureness held when the plan was first derived. When we receive knowledge of results of the actual outcome, it is possible to reconsider all of the sureness levels held throughout the process since derivation of the plan.

The main point of the above elaboration is that sureness about a predicted outcome is situational and subject to change due to the occurrence of events over time. While the daily life scheme described above clearly involves more time than a stimulus-response cycle of a paired-associates learning task, there is no apparent reason to assume that the CMO involved in the two cases are fundamentally different. Therefore, it can be inferred that sureness levels generated in the course of a PAL task are subject to change as a result of feedback over time. Model 1 (Appendix Table I-1) is a description of the ordinal sequence of sureness derivation and modification, as it is postulated to occur, in the normal learning condition of Hunt's PAL paradigm. This model represents the baseline process of covert assessment, as it is assumed to occur in all normal human operators. It should be clear from this model that the opportunity plainly exists within the S-R cycle of a PAL task for the generation and modification of covert assessments of sureness.

Model 2 (Appendix Table I-2) is a description of the ordinal sequences of sureness derivation and modification, as they are postulated to occur, in the MK and KM overt assessment treatments. Comparison of these two models shows that the MK and KM treatments have the same opportunity for sureness level modification that the normal learning treatment, M, does, plus an opportunity for sureness level modifcation as a

result of overt assessment execution feedback. Cross comparisons of the three treatments suggest that both MK and KM should enjoy some advantage over M, but other than the obvious order difference, nothing is suggested about the performance differential between MK and KM.

Discussion up to this point has attempted to describe the operational underpinnings of CMO, specifically the generation of covert assessments. Now, it is appropriate to consider the implications of a covert assessment process. The notion that the human operator employs a covert assessment mechanism to control the selection of responses for execution, suggests that generated assessments are salient factors in human performance.

Such salience might account for "why?" self-assessment facilitates learning, but it does not explain "how?". Available literature offers a couple of potential explainations of "how?". Hunt (1978) suggests, as one possibility, that the process of overt self-assessment increases covert assessment accuracy, which consequently causes the human operator to be better able to identify a correct response. An alternative possibility is suggested in an article by Fischoff, Slovic, and Lichtenstein (1977) in which it is reported that when people are asked to make overt confidence judgements they are consistently overconfident. The essence of this notion is that overt self-assessment distorts (inflates)covert assessment accuracy, which induces the human operator to overassess his/her ability to perform.

In operational terms of the PAL task the alternatives discussed above predict differing outcome propensities. There are four possible functional outcomes of assessment responding, which result from the fact that the subject will either be "sure" or "unsure" about his/her response, and that the response itself will be either correct or wrong. These four possible outcomes are sure-correct, sure-wrong, unsure-correct, and unsure-wrong. At the point in the PAL task when the subject receives knowledge of results, his/her expectations about a given outcome are either confirmed or disconfirmed. See Appendix Table I-3.

The assessment accuracy account of the effects of self-assessment predicts an increased proportion of confirmational outcomes of the types sure-correct and unsure-wrong, along with a decreased proportion of disconfirmational outcomes of the types sure-wrong and unsure-correct. The logic of enhanced assessment accuracy demands increased confirmation and decreased disconfirmation of outcomes.

The overconfidence account of the effects of self-assessment predicts an increased proportion of "sure" assessments and a decreased proportion of "unsure" assessments. This means an increase in the proportions of the confirmational outcome of sure-correct and the disconfirmational outcome of sure-wrong, as well as, a decrease in the proportions of the confirmational outcome of unsure-wrong and the disconfirmational outcome of unsure-correct.

In search of evidence that might indicate how overt self-assessment achieves its learning facilitation effect, whether it is through improved accuracy or through assessment inflation, this student carried out an item by item re-analysis of the data for all trials from the Hunt (1978) study. Ideally, this analysis would compare the covert assessment accuracies of the M, MK8, and K8M treatments, but since these are covert no such comparison is possible. Likewise, since the M treatment has no overt assessments, no comparison is possible between M and MK8 or K8M. The analysis had to be focused on the overt assessment differences between MK8 and K8M, with the premise that the findings would be consistent and parallel with the fact that MK8 required fewer trials to criterion than K8M and K8M required fewer than M.

The results of the analysis, shown in Table 2, tend to support the assessment inflation notion over the assessment accuracy notion. Items 1 and 2 in the table indicate that the MK8 group began using "sure" assessments (8 on a scale of 1 - 8) earlier in the PAL task, and used more of them than the K8M group. Additionally, the MK8 subjects were less accurate when they did use "sure" assessments, than were the K8M subjects, as shown in item number 3 of Table 2. Taken altogether, the information in items 1,2, and 3 of Table 2 describe an operational situation in which it was more likely that the MK8 subjects

Table 2

Comparison of Groups MK8 and K8M

Hunt's (1978) Data

Item	Category of Comparison	MK8	K8M
1	Mean proportion of extreme sure assessments.	.507	.432
2	Mean proportion of initial trials accomplished before the first extreme sure assessment was used.	.316	.472
3	Mean proportion of accurate extreme sure assessments. P(Correct∣Sure)	.836	.861
4	False Alarm Rate. P(Sure∣Wrong)	.237	.160
5	Hit Rate. P(Sure∣Correct)	.745	.621
6	Mean sureness level given correct answers.	7.03	6.82
7	Mean sureness level given wrong answers.	4.06	5.02

would experience greater confirmation and greater disconfirmation of "sure" assessments then the K8M subjects.

Taken alone, the three points of comparison discussed above are inconclusive. They describe the situation of greater likelihood of confirmation/disconfirmation for MK8 subjects, but they do not show that greater confirmation/disconfirmation was actually experienced. Item 4 of Table 2, false alarm rate, and item 5 of Table 2, hit rate, are metrics best suited for this purpose. False alarm rate (FA) is a conditional probability of a sure assessment given a wrong response. A comparison of the FA for the two groups suggests that the MK8 group did experience more disconfirmation of the type sure-wrong. The results for hit rate (HR), a conditional probability of a "sure" assessment given a correct response, are essentially the same. The MK8 group experienced more confirmation of the type sure-correct than did the K8M group. In terms of "sure" assessments, the MK8 group derived a situational advantage, due to greater assessment confidence, which allowed it to accrue whatever benefits there may be in increased expectation confirmation and/or disconfirmation.

The discussion of the analysis thus far has been concerned with the relative use of "sure" assessments. This approach is predicated on an observation by Hunt (1978) that his subjects treated the 1 - 8 assessment scale as being divided into two parts, with 1 - 7 signifying "unsure"

and 8 signifying "sure". Figure 1, which shows the conditional probabilities of a correct response given each of the eight levels of sureness on the 1 - 8 assessment scale, suggests that the subjects did use the scale in such a fashion. It is clearly the case for the MK8 group, but somewhat less so for the K8M group. Figure 1 seems to support the application of false alarm rate and hit rate as fairly valid comparative metrics. However, it should be noted that the two assessment orders tend to use the 1 - 8 assessment scale differently.

Items 6 and 7 of Table 2, which are previouly un-discussed, are the mean surenesses for the task as a whole given correct and wrong answers, respectively. Not surprisingly, the mean sureness level given correct answers is somewhat higher for the MK8 group than for the K8M group. However, item 7, the mean sureness level given wrong answers is somewhat lower for the MK8 group than for the K8M group. On surface this appears inconsistent with and contrary to the other results. Based on the assessment inflation notion and the results already discussed, the MK8 group would be expected to have a higher score here than K8M.

This seeming aberration in the data appears to be the result of differential assessment scale use. Figure 2 shows the percentage of use of each of the points on the 1 - 8 assessment scale for the two assessment orders. The upper portion of Figure 2 compares sureness level use given

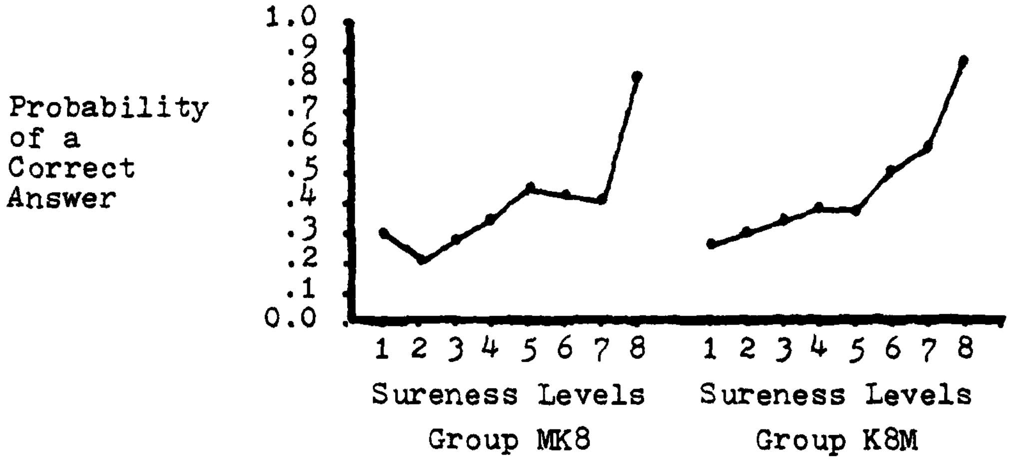
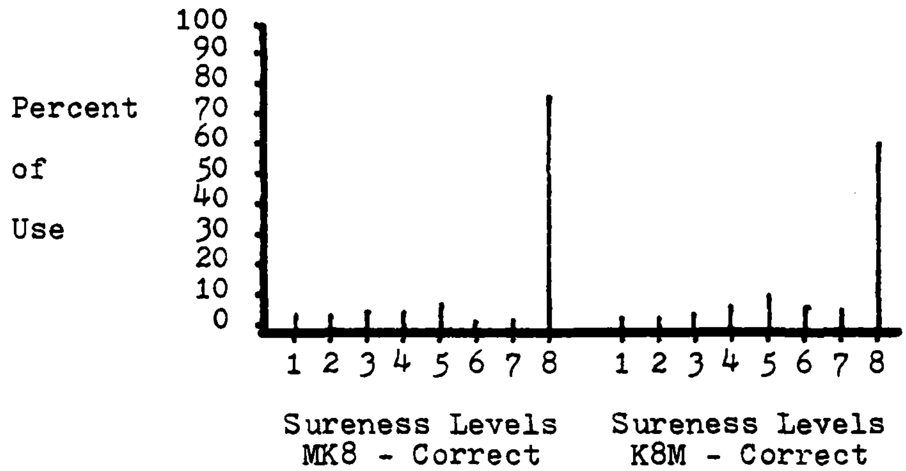


Figure 1. Conditional probability of a correct answer given each level of sureness on the 1 - 8 assessment scale for groups MK8 and K8M, using data from Hunt (1978).



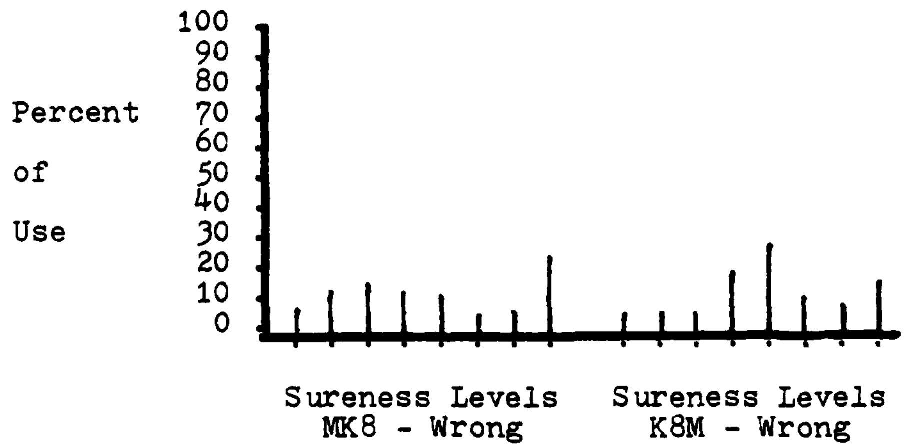


Figure 2. Percent of sureness level use given correct and wrong answers for groups MK8 and K8M, using data from Hunt (1978).

correct answers. For both the MK8 group and K8M group the modal sureness used is an 8, and as already discussed MK8 uses more of them. The lower portion compares percent of sureness level use given wrong answers. The modal sureness used by MK8 is again an 8, as would be expected. However, the modal sureness used by K8M is a 5, and the next most-used sureness is a 4. This predominant use of central sureness values by the K8M group serves to boost the mean sureness level given wrong answers, above that of the MK8 group. While this turn of events does emphasize differential scale use by the two assessment orders, it does not directly contradict the assessment inflation notion.

The analysis of Hunt's (1978) data discussed above seems to implicate overconfident assessment as a centrally important and potentially causal factor in self-assessment facilitation of learning. In order to further examine and attempt to verify this notion, additional study is needed.

The general replication of the procedures employed by Hunt (1978) is a basic step in the continuing study of self-assessment. However, additional manipulations are appropriate in order to better define the domain over which self-assessment might be effective. The study, herein reported, employs the essential procedures of the Hunt paradigm with four basic modifications concerning stimuli, stimulus sequence, levels of self-assessment, and nature of the self-assessment scale.

The first three of these modifications are fairly simple and straightforward. The present study uses consonant-vowel-consonant (CVC) trigrams with a 65% meaningfulness level (Archer, 1960) for stimuli, whereas the Hunt study used hand pliers as stimuli. This modification is cogent for purposes of identifying the types of stimuli materials the learning of which might be facilitated by means of self-assessment. The stimuli sequence for this study has been altered from that of Hunt's in an effort to better control for relative positional effects among the stimuli items. Lastly, this study uses only eight-point assessment scales, whereas the Hunt study used two-point, four-point, and eight-point scales.

The fourth modification is somewhat more complex and involves some theoretical aspects. In the discussion above it was pointed out that the two assessment orders, MK8 and K8M, lead subjects to use the 1 - 8 assessment scale differently. This fact, along with the results that show the MK8 group to be more confident in their assessments, suggest that the nature of the assessment scale might be of pivotal importance. Additionally, as Figure 1 shows, subjects tend to divide the 1 - 8 scale on the basis of 1 - 7, "unsure", and 8, "sure". It seems that this skewed scale division allows subjects to be ambiguous about both their assessments and their conviction in the correctness of their responses.

In an effort to counteract scale-use ambiguity and induce greater assessment conviction, a new, more precisely defined, eight-point scale is introduced in this study. Simply, this scale uses 4 negative numbers for "unsureness" and 4 positive numbers for "sureness", centered on an imagined zero mid-point. Extreme unsureness is represented by -4 and extreme sureness is represented by +4. Both Hunt's 1 - 8 scale and this new -4 thru +4 scale are used, so that a comparison can be drawn.

In this study, two variables, order of assessment and nature of assessment scale, are used in combination to define four experimental treatments. Additionally, a normal learning control condition and two motor control conditions are used for a total of seven conditions, as shown in Table 3.

There are two general hypotheses derived out of the foregoing data analysis and discussion. Each is presented below, in its turn, and immediately followed by a breakdown into its specifiable experimental hypotheses.

The first hypothesis is that the addition of an overt self-assessment step to the stimulus-response cycle of a paired-associates learning task will facilitate learning of CVC trigram pairs with a meaningfulness level of 65%.

a. Assessment treatments, MK8, K8M, MK$\pm$4, and K$\pm$4M, will require fewer trials to reach criterion, where criterion is the errorless completion of two

Table 3

Designs to Test Hypotheses One and Two

Hypothesis One

		Components of Response			
		Non-Assessment		Assessment	
		Learning Control	Motor Control	1 - 8	± 4
Order of Response	MK	M	MX	MK8	MK$\pm$4
	KM		XM	K8M	K$\pm$4M

Hypothesis Two

		Nature of Assessment Scale	
		1 - 8	± 4
Order of Assessment	MK	MK8	MK$\pm$4
	KM	K8M	K$\pm$4M

sequential trials, than the control conditions, M, MX, and XM, at the .05 level of significance.

b. Assessment treatments using the -4 thru +4 assessment scale, $MK\pm4$ and $K\pm4M$, will require fewer trials to reach criterion than the assessment treatments using the 1 - 8 scale, MK8 and K8M, respectively, at the .05 level of significance.

c. Assessment treatments using the MK order of responding, MK8 and $MK\pm4$, will require fewer trials to reach criterion than the assessment treatments using the KM order of responding, K8M and $K\pm4M$, respectively, at the .05 level of significance.

d. The $MK\pm4$ assessment treatment will require fewer trials to reach criterion than any other treatment, due to an over-additive interaction of the MK order of response and ±4 assessment scale.

The second hypothesis is that the process of overt self-assessment is associated with the overconfident use of "sure" assessments; with the MK order of assessment demonstrating greater overconfidence than the KM order of assessment, and with the -4 thru +4 assessment scale demonstrating greater overconfidence than the 1 - 8 assessment scale.

a. Assessment treatments using the -4 thru +4 assessment scale, $MK\pm4$ and $K\pm4M$, will have higher hit rates (HR), conditional probability of a "sure" assessment given a correct answer, and higher false alarm rates

(FA), conditional probability of a "sure" assessment given a wrong answer, than the assessment treatments using the 1 - 8 assessment scale, MK8 and K8M, respectively, at the .05 level of significance.

b. Assessment treatments using the MK order of responding, MK8 and MK$\pm$4, will have higher hit rates (HR) and higher false alarm rates (FA), than the assessment treatments using the KM order of responding, K8M and K$\pm$4M, respectively, at the .05 level of significance.

c. The MK$\pm$4 assessment treatment will have a higher hit rate (HR) and a higher false alarm rate (FA) than any other assessment treatment, due to an over-additive interaction interaction of the MK order of response and the $\pm$4 assessment scale.

Method

Subjects

The subjects used in this study were 140 university students enrolled in introductory psychology classes at New Mexico State University in the spring semester of 1980. The subjects participated in the experiment for course credit. Twenty subjects, ten female and ten male, were used in each of the seven conditions.

Apparatus

A Kodak carrousel slide projector, controlled by a PDP/8e DEC mini-computer, projected 35mm stimulus and knowledge-of-results slides onto a two-way screen. The projections were viewed by the subjects from inside a darkened and sound-shielded booth. The slides were seven duplicative sets of sixteen slides each. Eight slides from each set were stimulus CVC trigrams, shown as black letters on a white background. Each of these stimulus trigrams was immediately followed by a knowledge-of-results slide showing the same stimulus trigram paired with its response CVC trigram, also in black letters on a white background.

Sixteen CVC trigrams were randomly drawn from a list of trigrams with a 65% meaningfulness level, as classified by Archer (1960), and were randomly paired into the eight CVC trigram pairs shown in Appendix Table I-4. The truncated Latin Square, shown in Appendix Table I-5, was used to order the CVC trigrams within sets to control for positional effects across sets.

Subject responses were collected by the PDP/8e DEC mini-computer via a push button keyboard situated between the subject and the visual display screen. Response characteristics were recorded in hard copy by a teletype interfaced with the mini-computer.

Procedure

Three variables, response order, assessment, and nature of assessment scale, were used to produce the seven treatments, as previouly shown in Table 3. Treatment variations were achieved by manipulation of button arrangements on the response keyboard. Appendix Table I-6 shows the button formats for the seven different treatments.

Three different control conditions were employed. The M condition is intended to represent normal learning using single-step responding. The MX and XM treatments were included to control for the effects of the two step responses necessary in the assessment procedures. The four experimental treatments were derived from factorial combinations of response order and nature of assessment scale.

Experimentation was counter-balanced across treatment groups for day-of-week and time-of day. Subjects were solicited into treatment groups, so that each group had ten females and ten males. The 140 subjects were examined one at a time during one-hour periods at the rate of twenty per week during an eight-week period. Appendix Table I-7 shows the planned seven-week experimentation schedule.

Subjects lost due to "no-shows" or equipment malfunctions were made up during the eighth week at times and on days that matched the original times and days as closely as possible.

Subjects were instructed for the experiment in accordance with the standardized instructions shown in Appendices A thru H. Appendices A thru G are treatment-specific instructions for the seven different treatments. Appendix H is a set of common instructions read to all subjects after their particular treatment instructions had been read.

Results

The results of this experiment, as shown in Table 4, do not support either of the two major hypotheses. The results for hypothesis one are displayed in terms of the dependent variable, trials to criterion. The results for hypothesis two are displayed in terms of hit rate (HR) and false alarm rate (FA).

Appendix Table I-8, the outcome of Dunnett's tD test (Winer, 1971, page 201) for comparing all means with a control, and Appendix Table I-9, the outcome of an analysis of variance (Myers, 1979), excluding the control condition, M, are tests of hypothesis 1a. Contrary to prediction of that hypothesis, the assessment treatments do not differ significantly from the control condition, M, $F(6,139)<1.0$, $p>.25$, or from the motor control conditions, MX and XM, $F(1,114)<1.0$, $p>.25$. Reference to Table 4 shows that only two of the assessment treatments, K8M and K$\pm$4M, had mean numbers of trials to criterion, 10.2 and 9.35 respectively, that are lower than the control condition means. The means for the three control conditions M, MX, and XM are 10.3, 10.3, and 10.9, respectively. The mean number of trials to criterion for the other two assessment treatments, MK8 and MK$\pm$4, are both higher than those of the control conditions at 11.15 and 10.55, respectively.

Hypotheses 1b, 1c, and 1d were tested by a 2 by 2 analysis of variance (Myers, 1979) the results of which are

Table 4

Data Comparison for the Various Treatments in Terms of Trials to Criterion, Hit Rate, and False Alarm Rate

Treatment	Mean	Median	Mode	Range	Standard Deviation
		Trials to Criterion			
M	10.30	9.5	6&13	17	4.1434
MX	10.30	10	10	14	3.8265
XM	10.90	11	11	11	2.8078
MK8	11.15	12.5	13	12	3.2971
K8M	10.20	10	10	12	3.3182
MK4	10.55	10	10&12	13	3.5165
K4M	9.35	9	8	8	2.2070
		Hit Rate			
MK8	.7505	.7500	.6551	.3291	.1142
K8M	.7096	.7367	a	.4519	.1274
MK4	.7185	.7097	a	.3572	.0883
K4M	.6278	.6966	.8125	.8750	.2316
		False Alarm Rate			
MK8	.0683	.0310	0	.2631	.0859
K8M	.0736	.0308	0	.2307	.0810
MK4	.0657	.0426	0	.3000	.0796
K4M	.0336	0	0	.2000	.0533

a: No values were repeated

shown in Appendix Table I-10. Hypothesis 1b predicts a main effect of the independent variable of nature of assessment scale, such that the -4 thru +4 scale should produce better performance than the 1 - 8 assessment scale. Comparison of mean number of trials to criterion shows that MK$\pm$4 and K$\pm$4M (10.55 and 9.35) did have better performance than MK8 and K8M (11.55 and 10.2), respectively, but the differences are slight and not significant, $F(1,76) = 1.075$, $p > .25$.

Hypothesis 1c predicts a main effect of the independent variable of order of assessment, such that the MK order of responding would have better performance than the KM order. Comparison of mean number of trials to criterion shows an effect opposite in direction from that predicted, such that K8M and K$\pm$4M (10.2 and 9.35) had better performance than MK8 and MK$\pm$4 (11.15 and 10.55), respectively. However, the differences are once again not significant, $F(1,76)=2.363$, $p>.10$.

Hypothesis 1d predicts an interaction of the -4 thru +4 assessment scale and the MK order of responding, such that the MK$\pm$4 assessment treatment would have the best overall performance. As already observed above, this was not the case in this experiment. Both the K$\pm$4M and K8M conditions (9.35 and 10.2) showed better performance than the MK$\pm$4 condition (10.55). No significant interaction was observed, $F(1,76) < 1.0$, $p>.25$.

Appendix Tables I-11 and I-12 display the results for tests of hypothesis two. Appendix Table I-10 is the outcome of a 2 by 2 analysis of variance (Myers, 1979) using hit rate (HR) as the dependent variable, and Appendix Table I-11 is the same except that it uses false alarm rate (FA) as the dependent variable.

Hypothesis 2a predicts a main effect of the independent variable of <u>nature of assessment scale</u>, such that the -4 thru +4 scale would have higher hit rates and false alarm rates than the 1 - 8 scale. In terms of hit rate an effect in the opposite direction is the case, where hit rates for MK8 and K8M (.7505 and .7096) are both higher than those for MK$\pm$4 and K$\pm$4M (.7185 and .6278), respectively. In any case, no significant main effect was observed, $F(1,76)=2.869$, $p > .10$. The case for false alarm rate is the same, with FA for MK8 and K8M (.0683 and .0736) both being higher than the FA for MK$\pm$4 and K$\pm$4M (.0657 and .0336), respectively. These differences are insignificant, $F(1,76)=1.570$, $p > .10$.

Hypothesis 2b predicts a main effect of the independent variable of <u>order of responding</u>, such that the MK order would have higher hit rates and false alarm rates than the KM order. In terms of hit rate (HR) the predicted outcome is the case, such that MK8 and MK$\pm$4 (.7505 and .7185) have higher hit rates than K8M and K$\pm$4M (.7096 and .6278), respectively. Although these differences lie in the

predicted direction, they are too slight to achieve desired significance, $F(1,76)=3.837$, $p>.05$. In terms of false alarm rate, a reverse effect is observed within the 1 - 8 scale, where FA for MK8 (.0683) is lower than that of K8M (.0736). However, within the -4 thru +4 scale the effect is in the predicted direction, with FA for MK$\pm$4 (.0657) being higher than that of K$\pm$4M (.0336). No significance is observed, $F(1,76)<1.0$, $p>.25$.

Hypothesis 2c predicts an interaction of the -4 thru +4 scale and the MK order of responding, such that the MK$\pm$4 assessment treatment would have the highest hit rate and highest false alarm rate of all the assessment treatments. This predicted effect is not observed in the data. In terms of hit rate the MK$\pm$4 treatment (.7185) is exceeded by the MK8 treatment (.7505). No significant interaction is observed, $F(1,76)<1.0$, $p>.25$. For the false alarm rate the MK$\pm$4 treatment (.0657) is exceeded by both the MK8 treatment (.0683) and the K8M treatment (.0736). Once again, no significance is observed, $F(1,76)=1.210$, $p>.25$.

Discussion

Clearly, from the results discussed above it is not reasonable to infer that any type of facilitation effect was in operation in favor of the self-assessment conditions in this experiment. While the negative outcome herein reported does serve to restrict the potential domain of self-assessment effectiveness, it will not serve to disprove the theoretical notions about self-assessment procedures offered in the introduction. Additional research is needed to further isolate and define the pertinent characteristics of self-assessment procedures necessary to achieve facilitation. Empirically, what is known presently is that in the Hunt (1978) study a significant self-assessment effect was observed, and in the present study such was not the case.

As was discussed in the introduction, this study was intended as a partial replication of the Hunt (1978) study with four modifications: stimuli, stimulus sequence, number of levels of self-assessment, and nature of self-assessment scale. The last three of these can not reasonably be assumed to have led to the negative results of this study. The modification of the stimulus sequence was a minor adjustment to the overall procedure, and the fact that this study did not employ assessment conditions of two and four levels of assessment, as Hunt (1978) did, can have no direct effect on the results. Likewise, while the inclusion of the new -4 thru +4 assessment scale could produce negative results

for and within themselves, they could not have a direct effect that would produce negative results for and within the original 1 - 8 assessment scale conditions in this experiment.

By the above process of elimination, modification of stimuli from pliers used by Hunt to CVC trigrams used in this study is implicated as a possible causal factor in accounting for the negative results herein reported. Clearly, CVC trigrams are different from pliers, but in what ways do they differ that might explain the different outcomes of the two studies? An empirical source of comparison is the overall mean number of trials to criterion for the two stimulus types. In Hunt's study the overall mean number of trials to criterion in learning the association of pliers with their names was approximately 18. In the present study the overall mean number of trials to criterion for learning associations of CVC trigram pairs was 10.39. On the face of this evidence it would seem that the CVC trigrams at a 65% meaningfulness level are less difficult to learn than pliers.

The implication of the above is that there may be some kind of floor effect limitation on the self-assessment facilitation of learning. Since the self-assessment process requires additional cognitive activity and one additional step in each overt response, it might reasonably be expected to produce some degree of interference with learning in

addition to and in spite of whatever facilitation it might generate. Under conditions where stimuli are unambiguous or the task is easy, such that normal learning procedures produce acquisiition in relatively few trials, it might be that inherent interference effects associated with a multi-stepped process such as self-assessment serve to counteract its learning facilitation.

Another difference between the two studies should be considered. In the current study a vast majority of the subjects, after hearing the standardized instructions, were unsure as to how they were supposed to know which answer was the correct one, and how they were supposed to be sure about their responses when initially they knew nothing about the stimulus set. The standard clarification given was that initially they could not possibly know which response was correct, so that they should guess in the response button row, and since they were guessing, initially their sureness should be relatively low. Knowledge of the stimulus set and sureness would increase over the course of the task.

The apparent result of this clarification was an increased use of the bottom sureness of the sureness scales, regardless of the nature of the scale. The high frequency of 1 on the 1 - 8 assessment scale and the -4 on the -4 thru +4 assessment scale is clearly not the case for the Hunt study, as is seen by a comparison among Figures 2, 3, and 4.

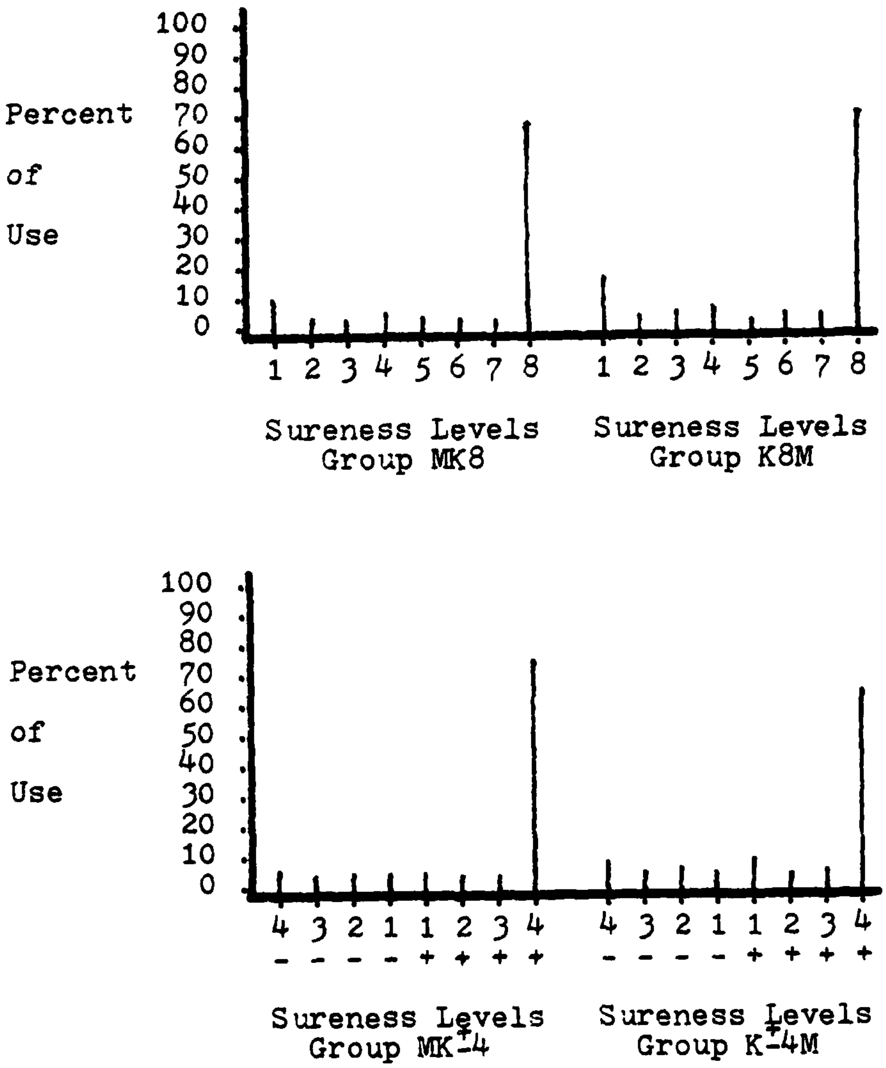


Figure 3. Percent of sureness level use given correct answers for groups MK8, K8M, MK±4, and K±4M.

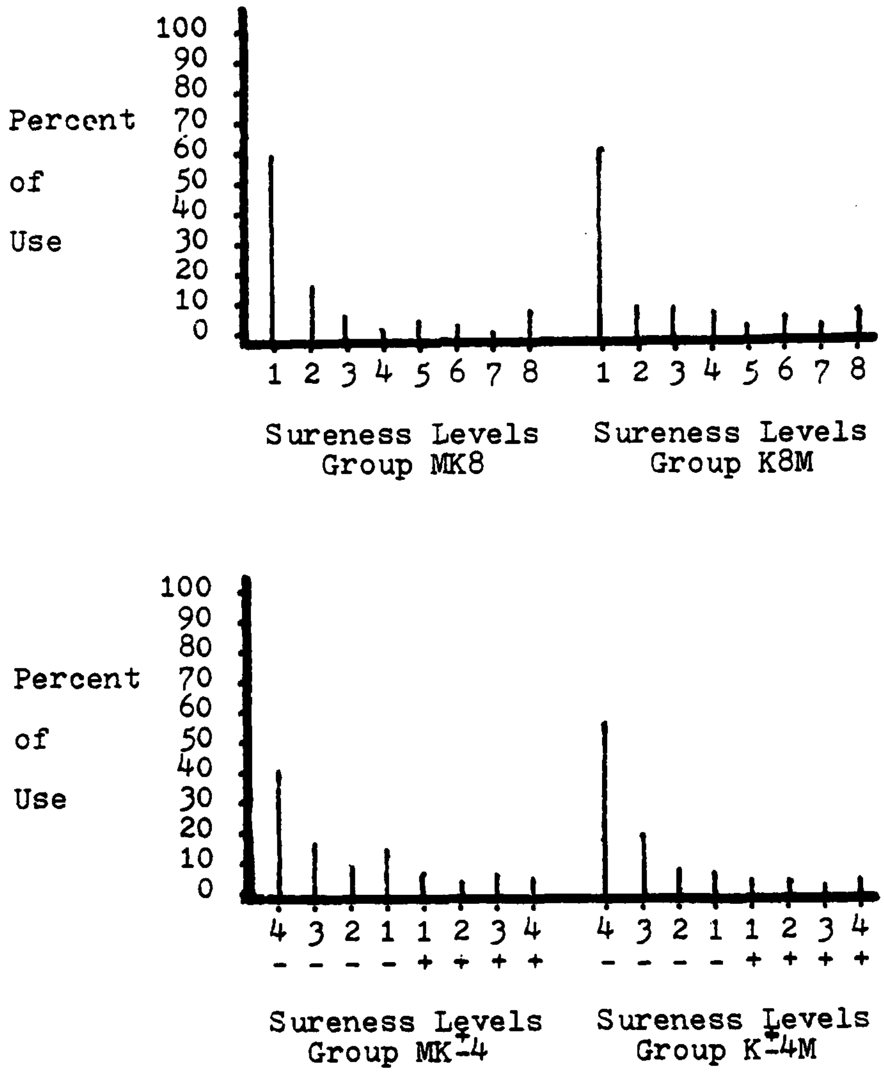


Figure 4. Percent of sureness level use given wrong answers for groups MK8, K8M, MK±4, and K±4M.

Figure 3 depicts the percentage of sureness level use given correct answers for all of the assessment groups. Comparison of Figure 2 with Figure 3 shows a slight increase in the use of low scale values in this study. Figure 4 depicts the percentage of sureness level use given wrong answers for all of the assessment groups. Comparison of Figure 2 with Figure 4 shows a marked increase in the use of low-scale values in this study.

A general trend observable in Figures 3, 4, and 5 is towards accuracy of assessment in this study. The probability curves for correctness given each level of the self-assessment scale, shown in Figure 5, tend to suggest that the subjects were reasonably accurate in their assessments of what they knew. This is more the case for the 1 - 8 assessment scale conditions where the curves are very regular, but appears to be generally the case across all four of the assessment conditions. Interestingly, the -4 thru +4 assessment scale conditions do seem to reflect the central division of the scale.

Comparison of the information in Table 5 for the Hunt study and this study appears to be consistent with the theoretical notions offered earlier and the differential outcomes of the two studies. Mean sureness given correct answers, mean sureness given wrong answers, and the conditonal probability of being wrong given extreme sureness are lower across all conditions of the current study

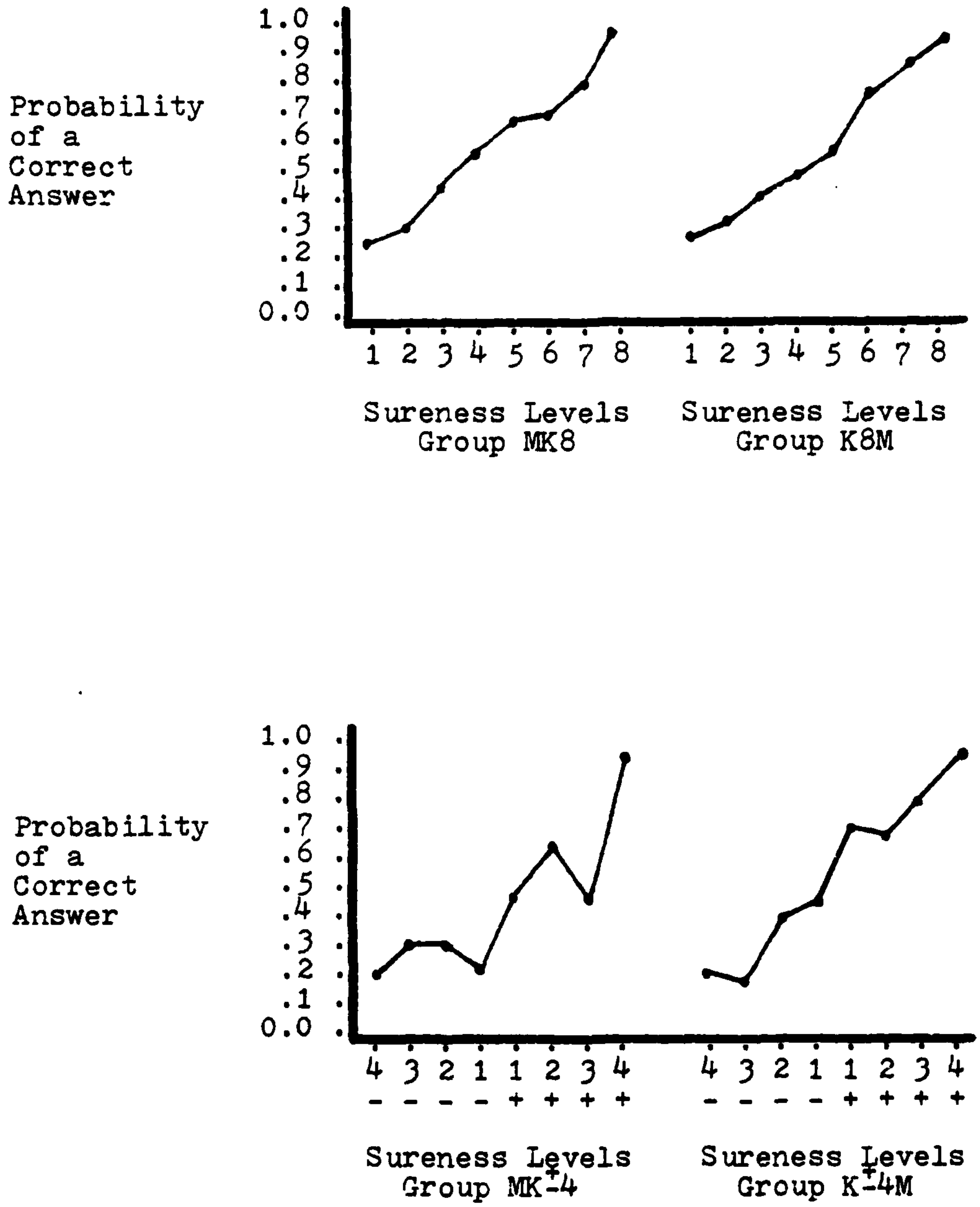


Figure 5. Conditional probability of a correct answer given each level of the sureness scale for groups MK8, K8M, MK±4, and K±4M.

Table 5

Comparison of Conditional Means and Probabilities

Category[a]	Current Study MK8	Current Study K8M	Current Study MK±4	Current Study K±4M	Hunt Study MK8	Hunt Study K8M
Mean($\bar{S} \mid C$)	6.56	6.53	6.77	6.55	7.03	6.82
Mean($\bar{S} \mid W$)	2.21	2.32	3.02	2.31	4.60	5.02
$P(W \mid S)$	.054	.065	.060	.035	.163	.138
$P(C \mid S)$	.945	.934	.939	.964	.836	.861
$P(S \mid W)$	.068	.073	.065	.033	.237	.160
$P(S \mid C)$	.750	.709	.718	.627	.745	.621

a: Key $\bar{S}$=Mean Sureness S=Extreme Sureness

W=Wrong C=Correct

than those of the Hunt study. Oppositely, the conditional probability of being correct given extreme sureness, an accuracy measure, is higher across all assessment treatments of the current study than it was for the Hunt study. The Hunt data is descriptive of a situation of high confidence, whereas the data for this study is descriptive of a situation of conservatism and accuracy. Given the theoretical framework herein espoused and the empirical information discussed just above, an *a priori* predictive process would predict better performance and more likely significance for the Hunt study relative to the current study.

The trend toward accuracy in this study could be the result of either the low ambiguity of the CVC trigrams or the clarification of the appropriate use of the assessment scale, or it could be due to both, separately or interactively. To maximize assessment effects in future research, stimuli should be realistically difficult and instructions for the use of the assessment scales should be minimal and possibly even ambiguous.

This last notion about ambiguity suggests that the inclusion of the -4 thru +4 scale expressly for the purpose of eliminating ambiguity might be an invalid approach. However, since the results of this study were negative, regardless of the nature of the assessment scale, the full effects of the -4 thru +4 assessment scale relative to the

1 - 8 scale are not known. Additional empirical evaluation of the two scales is appropriate before final conclusions are drawn.

Conclusion

In spite of the negative results observed in this study, the theoretical framework used to describe the self-assessment process receives inferential support from a cross-study comparison of the data and outcome of the Hunt study and the data and outcome of this study. The internal consistency of the theory with the observed empirical results is confirmed in pertinent literature.

According to Johnston (1979), the fundamental condition under which learning occurs is cortical arousal or desynchrony, which is reflected in large P_3 waveforms of the evoked potential. Events that have high utility (reinforcing) and high information (novel) have been shown to produce large P_3 waveforms by Sutton, Braren, Zubin, and John (1965).

Overconfidence in overt assessments might reasonably be expected to produce the novelty and cortical arousal necessary for learning. A response about which an individual is extremely sure, but that turns out to be wrong should produce "surprise", which is a novel event. Surprise would also be the result for the case where a "no confidence" assessment response turns out to be a correct answer. These two possibilities constitute the disconfirmational outcomes

discussed in the introduction: sure-wrong and unsure-correct. However, as was pointed out in that portion of the introduction, overconfident assessment should also lead to an increased rate of confirmation, as well as disconfirmation. Whereas disconfirmation is consistent with high information (novel) learning, confirmation appears to be consistent with high utility (reinforcing) learning. The learning facilitator that can accrue to itself the advantages of both greater confirmation and greater disconfirmation of expectancies, should also accrue more rapid learning and better performance.

In view of the positive results of the Hunt study and the theoretical consistency discussed above, continued research into the characteristics of self-assessment appears appropriate and justified. Assuming the eventual isolation of the circumstances critical to learning facilitation by means of self-assessment, the practical applications of self-assessment in learning and training are potentially broad and important from the training management standpoint.

Self-assessment procedures are relatively simple to employ, and the mechanics involved would probably become virtually second nature to users once they became familiar with them. Economically, the broad application of self-assessment procedures need not be prohibitive, since self-assessment is a procedure functionally dependent upon a set of events already available in situations involving learning and training, that is, cognition.

References

Archer, E.J. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs: General and Applied, 1960, 74(10), number 497, pages 1-23.

Attneave, F. How do you know? American Psychologist, 1974, 29, 493-499.

Fischoff, B., Slovic, P., and Lichtenstein, S. Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology, 1977, 3, 552-564.

Hunt, D.P. A Human Self-assessment Process. Final report to U.S. Army Research Institute for the Behavioral and Social Sciences, 1978.

Johnston, V.S. Stimuli with biological significance. In Beglighter, H. (Ed.), Evoked Brain Potentials and Behavior, New York: Plenum Press, 1979.

Kelly, C.R. Manual and Automatic Control. New York: John Wiley and Sons, 1968.

Miller, G.A., Galanter, E., and Pribram, K.H. Plans and the Structure of Behavior, New York: Henry Holt and Co., 1960.

Myers, J.L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, Inc., 1979.

Sutton, S., Braren, M., Zubin, J., and John, E.R. Evoked-potential correlates of stimulus uncertainty. Science, 1965, 150, 1187-1188.

Winer, B.J. Statistical Principles In Experimental Design New York: McGraw-Hill Book Co., 1971.

Appendix A

Treatment-Specific Portion of Group M Instructions

Instruct group M subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. The stimulus member of each pair will be presented on this screen. (Point) You will press one of the buttons in this response row to indicate the trigram that you believe is the correct response member of the pair. (Point to response button row sweepingly.) You should try to press the response button as quickly and as accurately as possible.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you <u>must</u> hold the "start button" down. (Point) A stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must release the "start button" and press the correct response button as fast and as accurately as possible. Next, the stimulus trigram will go off, and the correctly associated pair will reappear for a period of four seconds."

Appendix B

Treatment-Specific Portion of Group MX Instructions

Instruct group MX subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. The stimulus member of each pair will be presented on this screen. (Point) First, you will press one of the buttons in this response row to indicate the trigram that you believe is the correct response member of the pair, (Point to response button row sweepingly.) and then you will press this button to continue the slide sequence. You should try to press the buttons as quickly and as accurately as possible. Make sure that you press the buttons in the order that I indicated. Response first, and then the "next slide" button.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you <u>must</u> hold the "start button" down. (Point) A stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must release the "start button" and press the correct response button

and then the "next slide" button as fast and as accurately as possible. Next, the stimulus trigram will go off, and the correctly associated pair will reappear for a period of four seconds."

Appendix C

Treatment-Specific Portion of Group XM Instructions

Instruct group XM subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. The stimulus member of each pair will be presented on this screen. (Point) First, you will press this button to continue the slide sequence, and then you will press one of the buttons in this response row to indicate the trigram that you believe is the correct response member of the pair.(Point to response button row sweepingly.) You should try to press the buttons as quickly and as accurately as possible. Make sure that you press the buttons in the order that I indicated. The "next slide" button first, and then a response button.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you <u>must</u> hold the "start button" down. (Point) A stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must release the "start button" and press the "next slide" button

and then the correct response button as fast and as accurately as possible. Next, the stimulus trigram will go off, and the correctly associated pair will re-appear for a period of four seconds."

Appendix D

Treatment-Specific Portion of Group MK8 Instructions

Instruct group MK8 subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. An additional task is to indicate on a scale of 1 thru 8 how sure you are that your responses are correct, with 1 being "not sure" and 8 being "sure". The stimuls member of each pair will be presented on this screen. (Point) First, you will press one of the buttons in this response row to indicate the trigram that you believe is the correct response member of the pair, (Point to response button row sweepingly.) and then you will press one of the buttons in this sureness row to indicate how sure you are that the response you gave was correct. You should try to press the buttons as quickly and as accurately as possible. Make sure that you press the buttons in the order that I indicated. Response first, and then sureness.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you <u>must</u> hold the "start button" down. (Point) A

stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must reslease the "start button", and press the correct response button and then a sureness button as fast and as accurately as possible. Next, the stimuls trigram will go off, and the correctly associated pair will reappear for a period of four seconds."

Appendix E

Treatment-Specific Portion of Group K8M Instructions

Instruct group K8M subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. An additional task is to indicate on a scale of 1 thru 8 how sure you are that your responses will be correct, with 1 being "not sure" and 8 being "sure". The stimulus member of each pair will be presented on this screen. (Point) First, you will press one of the buttons in this sureness row to indicate how sure you are that the response you will give will be correct, (Point to sureness button row.) and then you will press one of the buttons in this response row to indicate the trigram that you believe is the correct response member of the pair. You should try to press the buttons as quickly and as accurately as possible. Make sure that you press the buttons in the order that I indicated. Sureness first, and then response.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you <u>must</u> hold the "start button" down. (Point) A

stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must release the "start button", and press a sureness button and then the correct response button as fast and as accurately as possible. Next, the stimulus trigram will go off, and the correctly associated pair will reappear for a period of four seconds."

Appendix F

Treatment-Specific Portion of Group $MK^{\pm}4$ Instructions

Instruct group $MK^{\pm}4$ subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. An additional task is to indicate on a scale of -4 thru +4 how sure you are that your responses are correct, with negative numbers being degrees of "unsureness" and positive numbers being degrees of "sureness". Minus 4 represents extreme unsureness, and plus 4 represents extreme sureness. The stimulus member of each trigram pair will be presented on this screen. (Point) First, you will press one of the buttons in this response row to indicate the trigram that you believe is the correct response member of the pair, (Point to response button row sweepingly.) and then you will press one of the buttons in this sureness row to indicate how sure you are that the response you gave was correct. You should try to press the buttons as quickly and as accurately as possible. Make sure that you press the buttons in the order that I indicated. Response first, and then sureness.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will

hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you must hold the "start button" down. (Point) A stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must release the "start button", and press the correct response button and then a sureness button as fast and as accurately as possible. Next, the stimulus trigram will go off, and the correctly associated pair will reappear for a period of four seconds."

Appendix G

Treatment-Specific Portion of Group $K\pm4M$ Instructions

Instruct group $K\pm4M$ subjects using the following two paragraphs, and the common instructions in Appendix H.

"In this experiment your task is to learn to properly associate eight different nonsense trigram pairs. An additional task is to indicate on a scale of -4 thru +4 how sure you are that your responses will be correct, with negative numbers being degrees of "unsureness" and positive numbers being degrees of "sureness". Minus 4 represents extreme unsureness, and plus 4 represents extreme sureness. The stimulus member of each trigram pair will be presented on this screen. (Point) First, you will press one of the buttons in this sureness row to indicate how sure you are that the response you will give will be correct, (Point to sureness button row.) and then you will press one of the buttons in this response row to indicate the trigram you believe is the correct response member of the pair. You should try to press the buttons as quickly and as accurately as possible. Make sure that you press the buttons in the order that I indicated. Sureness first, and then response.

When we start the experiment, you will put on these earphones. (Point) During the experiment, each trial will take place in the following order. First, you will

hear a short tone through the earphones, which will inform you that a stimulus trigram is about to be presented. Then, you <u>must</u> hold the "start button" down. (Point) A stimulus trigram will appear on the screen and stay on for eight seconds. During the eight seconds, you must release the "start button", and press a sureness button and then the correct response button as fast and as accurately as possible. Next, the stimulus trigram will go off, and the correctly associated pair will reappear for a period of four seconds."

Appendix H

Instructions Commom to All Treatments

Instruct all subjects by using the treatment-specific instructions contained in Appendices A thru G first, and then the common instructions below.

"Immediately, the brief tone will sound again to signal that the next stimulus trigram is to be presented. Make sure that you press the "start button" when you hear the tone and that you hold it down until the trigram is presented. If you are not pressing the "start button", when the stimulus trigram is supposed to be presented a wrong response will be recorded and you will see a blank screen for a period of nine seconds.

In a moment we will start a practice session that will consist of two sets of eight pairs of slides, with each set preceded by a "dot" slide. I will assist you in becoming familiar with the sequence of the task as necessary. You must press a response button every time. This is very important. At first, it may seem that the trial sequence happens quite rapidly, and you may get out of sequence with it. You will know that this has happened if there are long periods with nothing on the screen. You can get back in sequence by pressing the "start button" when you hear the tone, and holding it down until a stimulus trigram is presented. Are there any questions

before we begin? (Clarify as necessary.) OK, please put on the earphones for the practice session."

AFTER THE PRACTICE SESSION:

"Do you have any questions about the sequence of events? (Clarify as necessary.) During the experiment, the eight trigram pairs will be presented in a different order for each trial, until you have gone thru the sequence twice without an error. When you have achieved this level of learning, the computer, which is controlling the experiment, will terminate the slide sequence, and I will notify you that the experiment is over. Do not stop making responses just because you believe that you have reached the required level of learning performance. Continue to make responses as long as trigrams are being presented. Do you have any final questions? (Clarify as necessary.)"

Appendix I

Supplementary Data Tables

Appendix Table I-1

Ordinal Sequence of Sureness Derivation and Modification

Normal Learning Conditon M

CMO Opportunity[a] $S - c_1 - R - c_2 - KR - c_3$

Derivation[a] $S - m - k - R - k_a - k_m - KR - k_c/k_w - k_e$

$\underline{c_1}$:

1) Select m from memory.
2) Model m to simulate its execution and effect.
3) Generate k, sureness about m.
4) Compare k to criterion k; if
   k ≥ criterion k: issue m, or if
   k < criterion k: select m_2 and repeat modeling.

$\underline{c_2}$:

1) Receive intrinsic feedback from response execution of M.
2) Compare actual execution, M, and expected execution, m.
3) Generate k_a, sureness about M.
4) Calculate k_m, response bias term.

$\underline{c_3}$:

1) Receive KR.
2) Compare S and R(M) with KR;
   if correct, generate appropriate k_c: P = 1.0, or
   if wrong, generate appropriate k_w: P = 0.0.
3) Compare appropriate k_c/k_w with k_a, and generate k_e, error term.
4) Audit sureness sequence, comparing k, k_a, k_c/k_w.
5) Adjust assessment mechanism in terms of k_m and k_e.

---

a: Key

Covert Events		Overt Events	
m:	Tentative response	R=M:	Actual Response
k:	Covert Sureness	S:	Stimulus
$c_{\#}$:	Opportunity for Cognition	KR:	Knowledge of Results
		K:	Overt Sureness

Appendix Table I-2

Ordinal Sequence of Sureness Derivation and Modification

MK and KM Assessment Conditions

CMO Opportunity:

MK: $S - c_1 - R - c_2 - K - c_3 - KR - c_4$
KM: $S - c_1 - K - c_2 - R - c_3 - KR - c_4$

Derivation:

MK: $S - m - k - R - k_a - k_m - K - k_A - k_k - KR - k_c/k_w - k_e$
KM: $S - m - k - K - k_A - k_k - R - k_a - k_m - KR - k_c/k_w - k_e$

c_1 for both MK and KM:

1) Select m from memory.
2) Model m to simulate its execution and effect.
3) Generate k, sureness about m.
4) Compare k with criterion k; if
   $k \geq$ criterion k: issue m, or if
   $k <$ criterion k: select m_2 and repeat modeling.

c_2 for MK and c_3 for KM:

1) Receive intrinsic feedback from response execution of M.
2) Compare actual execution, M, and expected execution, m.
3) Generate k_a, sureness about K.
4) Calculate k_m, response bias term.

c_3 for KM and c_2 for KM:

1) Receive intrinsic feedback from assessment execution of K.
2) Compare actual execution, K, and expected execution, k.
3) Generate k_A, sureness about K.
4) Calculate k_k, assessment bias term.

c_4 for both MK and KM:

1) Receive KR.
2) Compare S, R(M), and KR;
   if correct, generate appropriate k_c: P = 1.0, or
   if wrong, generate appropriate k_w: P = 0.0
3) Compare appropriate k_c/k_w with k_a or k_A, and generate k_e, error term.
4) Audit sureness sequence, comparing k, k_a, k_A, and k_c/k_w.
5) Adjust assessment mechanism in terms of k_m, k_k, and k_e.

Appendix Table I-3

Functional Outcomes of Assessment Responding And Response Proportion Prediction of Accuracy And Overconfidence Accounts of the Self-Assessment Effect

	Outcome Components			
	Sure		Unsure	
Issue	Correct	Wrong	Correct	Wrong
Type of Outcome	Sure-Correct	Sure-Wrong	Unsure-Correct	Unsure-Wrong
Category of Outcome	Confirm	Dis-Confirm	Dis-Confirm	Confirm
"Biological" Effect of Outcome	Rein-forcement	Novelty	Novelty	Rein-forcement
Prediction of Accuracy Account	$\wedge_a$	$\vee$	$\vee$	$\wedge$
Prediction of Overconfidence Account	$\wedge$	$\wedge$	$\vee$	$\vee$

a: Key $\wedge$: Increase in the proportion of the type of outcome over the course of the PAL task, or a higher rate of the type of outcome relative to that of a non-assessment or less effective assessment situation.

$\vee$: Decrease in the proportion described above.

Appendix Table I-4

Stimulus-Response CVC Trigram Pairs[a]

Number	Stimulus - Response
1	WOK - FON
2	HET - WID
3	LOD - SYP
4	KAF - HUD
5	MEY - VIX
6	NYC - KOC
7	VIR - KEL
8	PAB - LEP

a: All trigrams have association values of 65, as rated by Archer (1960).

Appendix Table I-5

Truncated Latin Square Assignment of Stimulus Items

		Item Number							
		1	2	3	4	5	6	7	8
	1	2	1	3	8	4	7	5	6
	2	3	2	4	1	5	8	6	7
Set	3	4	3	5	2	6	1	7	8
Number	4	5	4	6	3	7	2	8	1
	5	6	5	7	4	8	3	1	2
	6	7	6	8	5	1	4	2	3
	7	8	7	1	6	2	5	3	4

Appendix Table I-6

Button Arrangements for Treatment Variations

Treatments[a]	Button Arrangements
M	M M M M M M M M S
MX	X M M M M M M M M S
XM	M M M M M M M M X S
MK8	K K K K K K K K (1) (●) (●) (●) (●) (●) (●) (8) M M M M M M M M S
K8M	M M M M M M M M K K K K K K K K (1) (●) (●) (●) (●) (●) (●) (8) S
MK±4	K K K K K K K K (4) (3) (2) (1)-0+(1) (2) (3) (4) M M M M M M M M S
K±4M	M M M M M M M M K K K K K K K K (4) (3) (2) (1)-0+(1) (2) (3) (4) S

a: Key

M: Response
K: Assessment
X: Motor Component
S: Start Button
●: Unspecified point on 1 - 8 scale

Appendix Table I-7

Experimentation Schedule

Week	Time	Day-of-Week[a] M	T	W	Th	F
1	0830	M/M	K8M/F	K4M/M	MX/F	MK4/M
	0930	MK4/F	MX/M	M/F	K8M/M	M/F
	1030	K8M/M	XM/F	XM/M	XM/F	MK8/M
	1130	MX/F	MK4/M	MK8/F	MK8/M	K4M/F
2	0830	MX/F	MK4/M	M/F	XM/M	K4M/F
	0930	K4M/M	XM/F	MX/M	MK4/F	MX/M
	1030	MK4/F	MK8/M	MK8/F	MK8/M	K8M/F
	1130	XM/M	K4M/F	K8M/M	K8M/F	M/M
3	0830	XM/M	K4M/F	MX/M	MK8/F	M/M
	0930	M/F	MX/M	XM/F	K4M/M	XM/F
	1030	K4M/M	K8M/F	K8M/M	K8M/F	MK4/M
	1130	MK8/F	M/M	MK4/F	MK4/M	MX/F
4	0830	MK8/F	M/M	XM/F	K8M/M	MX/F
	0930	MX/M	K8M/F	MK8/M	M/F	MK8/M
	1030	M/F	MK4/M	MK4/F	MK4/M	K4M/F
	1130	K8M/M	MX/F	K4M/M	K4M/F	XM/M
5	0830	K8M/M	MX/F	MK8/M	MK4/F	XM/M
	0930	XM/F	K4M/M	K8M/F	MX/M	K8M/F
	1030	MX/M	K4M/F	K4M/M	K4M/F	M/M
	1130	MK4/F	XM/M	M/F	M/M	MK8/F
6	0830	MK4/F	XM/M	K8M/F	K4M/M	MK8/F
	0930	MK8/M	K4M/F	Mk4/M	XM/F	MK4/M
	1030	XM/F	M/M	M/F	M/M	MX/F
	1130	K4M/M	MK8/F	MX/M	MX/F	K8M/M
7	0830	K4M/M	MK8/F	MK4/M	M/F	K8M/M
	0930	K8M/F	M/M	K4M/F	MK8/M	K4M/F
	1030	MK8/M	XM/F	MX/M	MX/F	XM/M
	1130	M/F	K8M/M	MK8/F	XM/M	MK4/F

a: Treatment/Sex of subject

Appendix Table I-8

Outcome of Dunnett's tD Test

All Means Compared With a Control

Trials to Criterion

Overall F

Source of Variation	Df	SS	MS	F
Treatments	6	39.942	6.657	<1.0
Experimental Error	139	1496.450	10.765	
Total	145	1536.392		

Indivdiual Comparisons

Control	Comparison Condition	tD
M	MX	0
M	XM	.5782
M	MK8	.8192
M	K8M	-.0963
M	$MK^{\pm 4}$	.2409
M	$K^{\pm}4M$	-.9155

Appendix Table I-9

Outcome of Analysis of Variance

Excluding the Control Condition (M)

Trials to Criterion

Data Format

Nature of Assessment

Order of Response			
	MX	MK8	MK$\pm$4
	XM	K8M	K$\pm$4M

Summary

Source	Df	SS	MS	F
Nature	2	12.7166	6.3583	<1.0
Order	1	8.0083	8.0083	<1.0
Interaction	2	19.0166	9.5083	<1.0
Subjects	114	1171.2500	10.2741	
Total	119	1210.9916		

Appendix Table I-10

Outcome of Analysis of Variance

Assessment Conditions - Trials to Criterion

Data Format

Nature of Assessment

Order of Response	MK8	MK$\pm$4
	K8M	K$\pm$4M

Summary

Source	Df	SS	MS	F
Nature	1	10.5125	10.5225	1.075
Order	1	23.1125	23.1125	2.363
Interaction	1	.3125	.3125	$<$1.0
Subjects	76	743.2500	9.7796	
Total	79	777.1875		

Appendix Table I-11

Outcome of Analysis of Variance

Assessment Conditions - Hit Rate

Data Format

Nature of Assessment

Order of Response		
	MK8	$MK\pm4$
	K8M	$K\pm4M$

Summary

Source	Df	SS	MS	F
Nature	1	.0647	.0647	2.869
Order	1	.0866	.0866	3.837
Interaction	1	.0123	.0123	<1.0
Subjects	76	1.7163	.0225	
Total	79	1.8801		

Appendix Table I-12

Outcome of Analysis of Variance

Assessment Conditions - False Alarm Rate

Data Format

Nature of Assessment

Order of Response		
	MK8	MK$\pm$4
	K8M	K$\pm$4M

Summary

Source	Df	SS	MS	F
Nature	1	.009090	.009090	1.075
Order	1	.003592	.003592	2.363
Interaction	1	.007007	.007007	$<$1.0
Subjects	76	.440121	.005791	
Total	79	.459810		